

0570
/105

2

OIPe

RAW SEQUENCE LISTING

DATE: 01/26/2002

PATENT APPLICATION: US/09/981,947

TIME: 14:16:48

Input Set : N:\Cr3\RULE60\09981947.raw

Output Set: N:\CRF3\01252002\I981947.raw

SEQUENCE LISTING

1 (1) GENERAL INFORMATION:

2 (i) APPLICANT: Tartaglia, Louis A.
3 Weng, Xun

4 (ii) TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
5 GLUTEX AND USES THEREOF

6 (iii) NUMBER OF SEQUENCES: 10

7 (iv) CORRESPONDENCE ADDRESS:

8 (A) ADDRESSEE: Fish & Richardson P.C.

9 (B) STREET: 225 Franklin Street

10 (C) CITY: Boston

11 (D) STATE: MA

12 (E) COUNTRY: USA

13 (F) ZIP: 02110-2804

14 (v) COMPUTER READABLE FORM:

15 (A) MEDIUM TYPE: Diskette

16 (B) COMPUTER: IBM Compatible

17 (C) OPERATING SYSTEM: Windows95

18 (D) SOFTWARE: FastSEQ for Windows Version 2.0

19 (vi) CURRENT APPLICATION DATA:

C--> 20 (A) APPLICATION NUMBER: US/09/981,947

C--> 21 (B) FILING DATE: 18-Oct-2001

22 (vii) PRIOR APPLICATION DATA:

23 (A) APPLICATION NUMBER: 09/610,417

24 (B) FILING DATE:

25 (viii) ATTORNEY/AGENT INFORMATION:

26 (A) NAME: Meiklejohn, Ph.D., Anita L.

27 (B) REGISTRATION NUMBER: 35,283

28 (C) REFERENCE/DOCKET NUMBER: 07334/072002

29 (ix) TELECOMMUNICATION INFORMATION:

30 (A) TELEPHONE: 617/542-5070

31 (B) TELEFAX: 617/542-8906

32 (C) TELEX: 200154

33 (2) INFORMATION FOR SEQ ID NO: 1:

34 (i) SEQUENCE CHARACTERISTICS:

35 (A) LENGTH: 2343 base pairs

36 (B) TYPE: nucleic acid

37 (C) STRANDEDNESS: single

38 (D) TOPOLOGY: linear

39 (ii) MOLECULE TYPE: cDNA

40 (ix) FEATURE:

41 (A) NAME/KEY: Coding Sequence

42 (B) LOCATION: 73...1761

ENTERED

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/981,947

DATE: 01/26/2002

TIME: 14:16:48

Input Set : N:\Crf3\RULE60\09981947.raw

Output Set: N:\CRF3\01252002\I981947.raw

```

43 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
44 TCGACCCACG CGTCCGGCCT TGGCAGAGTC TGGGGTCCCT GGACTGAGCC ATCAGCTGGG 60
45 TCACTGAGAC CC ATG GCA AGG AAA CAA AAT AGG AAT TCC AAG GAA CTG GGC 111
46 Met Ala Arg Lys Gln Asn Arg Asn Ser Lys Glu Leu Gly
47 1 5 10
48 CTA GTT CCC CTC ACA GAT GAC ACC AGC CAC GCC GGG CCT CCA GGG CCA 159
49 Leu Val Pro Leu Thr Asp Asp Thr Ser His Ala Gly Pro Pro Gly Pro
50 15 20 25
51 GGG AGG GCA CTG CTG GAG TGT GAC CAC CTG AGG AGT GGG GTG CCA GGT 207
52 Gly Arg Ala Leu Leu Glu Cys Asp His Leu Arg Ser Gly Val Pro Gly
53 30 35 40 45
54 GGA AGG AGA AGA AAG GAC TGG TCC TGC TCG CTC CTC GTG GCC TCC CTC 255
55 Gly Arg Arg Arg Lys Asp Trp Ser Cys Ser Leu Leu Val Ala Ser Leu
56 50 55 60
57 GCG GGC GCC TTC GGC TCC TCC TTC CTC TAC GGC TAC AAC CTG TCG GTG 303
58 Ala Gly Ala Phe Gly Ser Ser Phe Leu Tyr Gly Tyr Asn Leu Ser Val
59 65 70 75
60 GTG AAT GCC CCC ACC CCG TAC ATC AAG GCC TTT TAC AAT GAG TCA TGG 351
61 Val Asn Ala Pro Thr Pro Tyr Ile Lys Ala Phe Tyr Asn Glu Ser Trp
62 80 85 90
63 GAA AGA AGG CAT GGA CGT CCA ATA GAC CCA GAC ACT CTG ACT CTG CTC 399
64 Glu Arg Arg His Gly Arg Pro Ile Asp Pro Asp Thr Leu Thr Leu Leu
65 95 100 105
66 TGG TCT GTG ACT GTG TCC ATA TTC GCC ATC GGT GGA CTT GTG GGG ACG 447
67 Trp Ser Val Thr Val Ser Ile Phe Ala Ile Gly Gly Leu Val Gly Thr
68 110 115 120 125
69 TTA ATT GTG AAG ATG ATT GGA AAG GTT CTT GGG AGG AAG CAC ACT TTG 495
70 Leu Ile Val Lys Met Ile Gly Lys Val Leu Gly Arg Lys His Thr Leu
71 130 135 140
72 CTG GCC AAT AAT GGG TTT GCA ATT TCT GCT GCA TTG CTG ATG GCC TGC 543
73 Leu Ala Asn Asn Gly Phe Ala Ile Ser Ala Ala Leu Leu Met Ala Cys
74 145 150 155
75 TCG CTC CAG GCA GGA GCC TTT GAA ATG CTC ATT GTG GGA CGC TTC ATC 591
76 Ser Leu Gln Ala Gly Ala Phe Glu Met Leu Ile Val Gly Arg Phe Ile
77 160 165 170
78 ATG GGC ATA GAT GGA GGC GTC GCC CTC AGT GTG CTC CCC ATG TAC CTC 639
79 Met Gly Ile Asp Gly Gly Val Ala Leu Ser Val Leu Pro Met Tyr Leu
80 175 180 185
81 AGT GAG ATC TCA CCC AAG GAG ATC CGT GGC TCT CTG GGG CAG GTG ACT 687
82 Ser Glu Ile Ser Pro Lys Glu Ile Arg Gly Ser Leu Gly Gln Val Thr
83 190 195 200 205
84 GCC ATC TTT ATC TGC ATT GGC GTG TTC ACT GGG CAG CTT CTG GGC CTG 735
85 Ala Ile Phe Ile Cys Ile Gly Val Phe Thr Gly Gln Leu Leu Gly Leu
86 210 215 220
87 CCC GAG CTG CTG GGA AAG GAG AGT ACC TGG CCA TAC CTG TTT GGA GTG 783
88 Pro Glu Leu Leu Gly Lys Glu Ser Thr Trp Pro Tyr Leu Phe Gly Val
89 225 230 235
90 ATT GTG GTC CCT GCC GTT GTC CAG CTG CTG AGC CTT CCC TTT CTC CCG 831
91 Ile Val Val Pro Ala Val Val Gln Leu Leu Ser Leu Pro Phe Leu Pro

```

RAW SEQUENCE LISTING

DATE: 01/26/2002

PATENT APPLICATION: US/09/981,947

TIME: 14:16:48

Input Set : N:\Crf3\RULE60\09981947.raw

Output Set: N:\CRF3\01252002\I981947.raw

92	240	245	250	
93	GAC AGC CCA CGC TAC CTG CTC TTG GAG AAG CAC AAC GAG GCA AGA GCT	879		
94	Asp Ser Pro Arg Tyr Leu Leu Leu Glu Lys His Asn Glu Ala Arg Ala			
95	255 260 265			
96	GTG AAA GCC TTC CAA ACG TTC TTG GGT AAA GCA GAC GTT TCC CAA GAG	927		
97	Val Lys Ala Phe Gln Thr Phe Leu Gly Lys Ala Asp Val Ser Gln Glu			
98	270 275 280 285			
99	GTA GAG GAG GTC CTG GCT GAG AGC CAC GTG CAG AGG AGC ATC CGC CTG	975		
100	Val Glu Glu Val Leu Ala Glu Ser His Val Gln Arg Ser Ile Arg Leu			
101	290 295 300			
102	GTG TCC GTG CTG GAG CTG CTG AGA GCT CCC TAC GTC CGC TGG CAG GTG	1023		
103	Val Ser Val Leu Glu Leu Leu Arg Ala Pro Tyr Val Arg Trp Gln Val			
104	305 310 315			
105	GTC ACC GTG ATT GTC ACC ATG GCC TGC TAC CAG CTC TGT GGC CTC AAT	1071		
106	Val Thr Val Ile Val Thr Met Ala Cys Tyr Gln Leu Cys Gly Leu Asn			
107	320 325 330			
108	GCA ATT TGG TTC TAT ACC AAC AGC ATC TTT GGA AAA GCT GGG ATC CCT	1119		
109	Ala Ile Trp Phe Tyr Thr Asn Ser Ile Phe Gly Lys Ala Gly Ile Pro			
110	335 340 345			
111	CCG GCA AAG ATC CCA TAC GTC ACC TTG AGT ACA GGG GGC ATC GAG ACT	1167		
112	Pro Ala Lys Ile Pro Tyr Val Thr Leu Ser Thr Gly Gly Ile Glu Thr			
113	350 355 360 365			
114	TTG GCT GCC GTC TTC TCT GGT TTG GTC ATT GAG CAC CTG GGA CGG AGA	1215		
115	Leu Ala Ala Val Phe Ser Gly Leu Val Ile Glu His Leu Gly Arg Arg			
116	370 375 380			
117	CCC CTC CTC ATT GGT GGC TTT GGG CTC ATG GGC CTC TTC TTT GGG ACC	1263		
118	Pro Leu Leu Ile Gly Gly Phe Gly Leu Met Gly Leu Phe Gly Thr			
119	385 390 395			
120	CTC ACC ATC ACG CTG ACC CTG CAG GAC CAC GCC CCC TGG GTC CCC TAC	1311		
121	Leu Thr Ile Thr Leu Thr Leu Gln Asp His Ala Pro Trp Val Pro Tyr			
122	400 405 410			
123	CTG AGT ATC GTG GGC ATT CTG GCC ATC ATC GCC TCT TTC TGC AGT GGG	1359		
124	Leu Ser Ile Val Gly Ile Leu Ala Ile Ile Ala Ser Phe Cys Ser Gly			
125	415 420 425			
126	CCA GGT GGC ATC CCG TTC ATC TTG ACT GGT GAG TTC TTC CAG CAA TCT	1407		
127	Pro Gly Gly Ile Pro Phe Ile Leu Thr Gly Glu Phe Phe Gln Gln Ser			
128	430 435 440 445			
129	CAG CGG CCG GCT GCC TTC ATC ATT GCA GGC ACC GTC AAC TGG CTC TCC	1455		
130	Gln Arg Pro Ala Ala Phe Ile Ile Ala Gly Thr Val Asn Trp Leu Ser			
131	450 455 460			
132	AAC TTT GCT GTT GGG CTC CTC TTC CCA TTC ATT CAG AAA AGT CTG GAC	1503		
133	Asn Phe Ala Val Gly Leu Leu Phe Pro Phe Ile Gln Lys Ser Leu Asp			
134	465 470 475			
135	ACC TAC TGT TTC CTA GTC TTT GCT ACA ATT TGT ATC ACA GGT GCT ATC	1551		
136	Thr Tyr Cys Phe Leu Val Phe Ala Thr Ile Cys Ile Thr Gly Ala Ile			
137	480 485 490			
138	TAC CTG TAT TTT GTG CTG CCT GAG ACC AAA AAC AGA ACC TAT GCA GAA	1599		
139	Tyr Leu Tyr Phe Val Leu Pro Glu Thr Lys Asn Arg Thr Tyr Ala Glu			
140	495 500 505			

RAW SEQUENCE LISTING

DATE: 01/26/2002

PATENT APPLICATION: US/09/981,947

TIME: 14:16:48

Input Set : N:\Cr3\RULE60\09981947.raw

Output Set: N:\CRF3\01252002\I981947.raw

```

141 ATC AGC CAG GCA TTT TCC AAA AGG AAC AAA GCA TAC CCA CCA GAA GAG 1647
142 Ile Ser Gln Ala Phe Ser Lys Arg Asn Lys Ala Tyr Pro Pro Glu Glu
143 510 515 520 525
144 AAA ATC GAC TCA GCT GTC ACT GAT GCT CCT GCT TCT TCT CCT TTC ACT 1695
145 Lys Ile Asp Ser Ala Val Thr Asp Ala Pro Ala Ser Ser Pro Phe Thr
146 530 535 540
147 ACT CCG AAT ACA GCC TGG ATT CAA GCT GCC GCC ACC ACC ACC GCC ACC 1743
148 Thr Pro Asn Thr Ala Trp Ile Gln Ala Ala Thr Thr Thr Ala Thr
149 545 550 555
150 AAA AAA GAA CAC CCA TTG TAAACGGTCA TGTGGTATTT CCTCAACCTG GAATGACC 1799
151 Lys Lys Glu His Pro Leu
152 560
153 TTCCCTATC TTCTTCTCCT GGAGAACACC AAGTCATGAT GTCAGACAAG AGCTTGGATT 1859
154 TTGGAGACAT GGGTTTGAAT TCCAGTCATT CATTCTTTTA TTCAGCAAAT ATTTAACAAG 1919
155 TACTGACATG TCCCATATGT TGTTTTACCC ACTGGTTATA CAATGGGAGG GAGAGAGAGA 1979
156 GAGAGAGAGA GAGAGAGATG CTATTCTAAA AGCTTGAAGT CTAGGCTGTG CACGGTGGCT 2039
157 CACGCTGTA ATCCAGCAC TTTGGGAGGC CGAGGTGGGT GGATCGTGAG GTCAGGAGAT 2099
158 TGAGACCATC CTGGCTAACA TGGTGAACT CCCTCTCTAC TAAAAATACA AAAAATTAGC 2159
159 TGAGCATGGT GCGGGCGGCC TGTAAGTCCCA GCTACTTGGG AGGCTGAGGC AGGAGAATGG 2219
160 CGTGAACCCA GGAGGCGGAG CTTGCAGTGA GCCGAGATCA CACCACCACA CTCCAGCCTG 2279
161 GGTGACAGAG CCAGACTCCG TCTCAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAGGGCGG 2339
162 CCGC 2343

```

164 (2) INFORMATION FOR SEQ ID NO: 2:

165 (i) SEQUENCE CHARACTERISTICS:

166 (A) LENGTH: 563 amino acids

167 (B) TYPE: amino acid

168 (D) TOPOLOGY: linear

169 (ii) MOLECULE TYPE: protein

170 (v) FRAGMENT TYPE: internal

171 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```

172 Met Ala Arg Lys Gln Asn Arg Asn Ser Lys Glu Leu Gly Leu Val Pro
173 1 5 10 15
174 Leu Thr Asp Asp Thr Ser His Ala Gly Pro Pro Gly Pro Gly Arg Ala
175 20 25 30
176 Leu Leu Glu Cys Asp His Leu Arg Ser Gly Val Pro Gly Gly Arg Arg
177 35 40 45
178 Arg Lys Asp Trp Ser Cys Ser Leu Leu Val Ala Ser Leu Ala Gly Ala
179 50 55 60
180 Phe Gly Ser Ser Phe Leu Tyr Gly Tyr Asn Leu Ser Val Val Asn Ala
181 65 70 75 80
182 Pro Thr Pro Tyr Ile Lys Ala Phe Tyr Asn Glu Ser Trp Glu Arg Arg
183 85 90 95
184 His Gly Arg Pro Ile Asp Pro Asp Thr Leu Thr Leu Leu Trp Ser Val
185 100 105 110
186 Thr Val Ser Ile Phe Ala Ile Gly Leu Val Gly Thr Leu Ile Val
187 115 120 125
188 Lys Met Ile Gly Lys Val Leu Gly Arg Lys His Thr Leu Leu Ala Asn
189 130 135 140
190 Asn Gly Phe Ala Ile Ser Ala Ala Leu Leu Met Ala Cys Ser Leu Gln

```

RAW SEQUENCE LISTING

DATE: 01/26/2002

PATENT APPLICATION: US/09/981,947

TIME: 14:16:48

Input Set : N:\Crf3\RULE60\09981947.raw

Output Set: N:\CRF3\01252002\I981947.raw

191	145	150	155	160
192	Ala Gly Ala Phe Glu Met Leu Ile Val Gly Arg Phe Ile Met Gly Ile			
193		165	170	175
194	Asp Gly Gly Val Ala Leu Ser Val Leu Pro Met Tyr Leu Ser Glu Ile			
195		180	185	190
196	Ser Pro Lys Glu Ile Arg Gly Ser Leu Gly Gln Val Thr Ala Ile Phe			
197		195	200	205
198	Ile Cys Ile Gly Val Phe Thr Gly Gln Leu Leu Gly Leu Pro Glu Leu			
199		210	215	220
200	Leu Gly Lys Glu Ser Thr Trp Pro Tyr Leu Phe Gly Val Ile Val Val			
201		225	230	235
202	Pro Ala Val Val Gln Leu Leu Ser Leu Pro Phe Leu Pro Asp Ser Pro			
203		245	250	255
204	Arg Tyr Leu Leu Leu Glu Lys His Asn Glu Ala Arg Ala Val Lys Ala			
205		260	265	270
206	Phe Gln Thr Phe Leu Gly Lys Ala Asp Val Ser Gln Glu Val Glu Glu			
207		275	280	285
208	Val Leu Ala Glu Ser His Val Gln Arg Ser Ile Arg Leu Val Ser Val			
209		290	295	300
210	Leu Glu Leu Leu Arg Ala Pro Tyr Val Arg Trp Gln Val Val Thr Val			
211		305	310	315
212	Ile Val Thr Met Ala Cys Tyr Gln Leu Cys Gly Leu Asn Ala Ile Trp			
213		325	330	335
214	Phe Tyr Thr Asn Ser Ile Phe Gly Lys Ala Gly Ile Pro Pro Ala Lys			
215		340	345	350
216	Ile Pro Tyr Val Thr Leu Ser Thr Gly Gly Ile Glu Thr Leu Ala Ala			
217		355	360	365
218	Val Phe Ser Gly Leu Val Ile Glu His Leu Gly Arg Arg Pro Leu Leu			
219		370	375	380
220	Ile Gly Gly Phe Gly Leu Met Gly Leu Phe Phe Gly Thr Leu Thr Ile			
221		385	390	395
222	Thr Leu Thr Leu Gln Asp His Ala Pro Trp Val Pro Tyr Leu Ser Ile			
223		405	410	415
224	Val Gly Ile Leu Ala Ile Ile Ala Ser Phe Cys Ser Gly Pro Gly Gly			
225		420	425	430
226	Ile Pro Phe Ile Leu Thr Gly Glu Phe Phe Gln Gln Ser Gln Arg Pro			
227		435	440	445
228	Ala Ala Phe Ile Ile Ala Gly Thr Val Asn Trp Leu Ser Asn Phe Ala			
229		450	455	460
230	Val Gly Leu Leu Phe Pro Phe Ile Gln Lys Ser Leu Asp Thr Tyr Cys			
231		465	470	475
232	Phe Leu Val Phe Ala Thr Ile Cys Ile Thr Gly Ala Ile Tyr Leu Tyr			
233		485	490	495
234	Phe Val Leu Pro Glu Thr Lys Asn Arg Thr Tyr Ala Glu Ile Ser Gln			
235		500	505	510
236	Ala Phe Ser Lys Arg Asn Lys Ala Tyr Pro Pro Glu Glu Lys Ile Asp			
237		515	520	525
238	Ser Ala Val Thr Asp Ala Pro Ala Ser Ser Pro Phe Thr Thr Pro Asn			
239		530	535	540

VERIFICATION SUMMARY

DATE: 01/26/2002

PATENT APPLICATION: US/09/981,947

TIME: 14:16:49

Input Set : N:\Crf3\RULE60\09981947.raw

Output Set: N:\CRF3\01252002\I981947.raw

L:20 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:21 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
L:596 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=8
L:606 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=9
L:617 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:623 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:627 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:631 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:647 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:649 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:675 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:677 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10